DOCUMENT RESUME

ED 100 366 IR 001 465

AUTHOR	Diegert, Carl
TITLE	Some Experience with Interactive Computing in Teaching Introductory Statistics.
INSTITUTION	Cornell Univ., Ithaca, N.Y. Center for Environmental Quality Management.
PUB DATE	2 May 74
NOTE	15p.; Paper presented at the Annual Meeting of the Shared Educational Computing (2nd, New Paltz, New York, May 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Calculation; *College Students; *Computer Assisted Instruction; *Computer Programs; Computer Science; Statistical Analysis; Statistics
IDENTIFIERS	APL; Biostatistics; Computer Software; Cornell University; Statistical Packages

ABSTRACT

Students in two biostatistics courses at the Cornell Medical College and in a course in applications of computer science given in Cornell's School of Industrial Engineering were given access to an interactive package of computer programs enabling them to perform statistical analysis without the burden of hand computation. After a general discussion of the possible educational impact of the package, a brief report is given of its use in the above mentioned courses at Cornell. (Author)

ED100366

SOME EXPERIENCE WITH INTERACTIVE COMPUTING IN TEACHING INTRODUCTORY STATISTICS

by

Carl Diegert

Presented at the Shared
Educational Computing System
2nd Annual Meeting and Conference

May 2, 1974

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Center for Environmental
Quality Management
Cornell University
Ithaca, New York 14850

607-256-7323

IR 001 465

Some Experience with Interactive Computing in Teaching Introductory Statistics*

by

Carl Diegert
Center for Environmental
Quality Management,
Cornell University

Background

The drudgery of hand calculating and memorizing formulas typically dominates the experience of students in statistics courses. It is unfortunate that the development of computational skills has so often obscured conceptual understanding of statistical inference. With this in mind a group at Cornell prepared a package of interactive computer programs designed to substitute for some of the students' hand computation.

The package does not in itself teach statistics, and the manual describing it offers no specific instruction in statistical analysis. Students are expected to have some familiarity with the analyses they will attempt, either from a text or course lectures. Using the package requires virtually no knowledge of computing.[1] The user's manual fully describes the techniques

---

* I wish to thank Professors Miké and Severance, and Ms. Mushinski, for making available to me evaluations of the courses described in this paper.

required for communicating with the computer, and gives outlined steps and example solutions with comments. The interactive analyses are highly conversational and intended to be self-explanatory.

The package was written to be used within the framework of an already crowded course syllabus. However, it can be used effectively in large or small doses, depending upon the needs of the instructor. Students will benefit from dealing with large and complex problems, regardless of how many or how varied. Analysis of such problems can demonstrate the utility of formal statistical methods. By contrast, analysis of small problems need not have this result, for solutions here can be arrived at intuitively.

Why use the computer package?

The following seem to be the most important advantages of using the package. Depending on the extent to which it is used, it can play a small but useful role, or affect the entire curriculum and format of a statistics course.

Assigned Problems

1) A wide range of realistic and "relevant" problems can be assigned which, without the package, would require reduction of too much data, or unmanageably complex calculations. In a large class one could, for example, test the hypothesis that birthdays are distributed uniformly over the 365 days of the year.

2) Many problems can be assigned. In a survey course an instructor can require students to use a variety of analyses, without having to set aside class time to teach in detail the computational procedures involved. For example, the t-test for population mean could be described qualitatively as an extension of the test used when population variance is known.

3) Multiple problems, each involving the same analysis, can also be assigned. In the test-of-means example, one could "discover" that for a given sample, the confidence interval estimation is more accurate when the population variance is known. Repeated analysis of the same data could also demonstrate the effect of class interval definition on a frequency table and its grouped statistics.

Critical judgment

The students are assured a computationally correct analysis. Therefore the instructor can place emphasis upon their: 1) selecting the method of analysis appropriate to the problem at hand; 2) obtaining the relevant data and entering them correctly; and 3) giving a proper interpretation of the analysis results. For example, the problem might be posed of whether the males in the class demonstrated intellectual superiority over the females in their performance on the last examination. The students will then decide on an analysis and, perhaps, be asked to justify their selection. Should they choose a chi-square test of independence, their results might look similar to Figure 1A. Using different

assumptions, they might choose to test for a significant difference between sample means, and thus obtain results similar to Figure 1B. (See next page for Figure 1.)

Their interpretation of the output is crucial, and will reveal the extent to which they understand the use of the analysis. It will also show whether they grasp the general and important point that different kinds of analysis can be made of the same problem, and that they can yield rather different results.

Learning from experience

Students using the package can learn from their mistakes, something that rarely occurs when they study worked problems. For example, suppose students are to learn the use of regression analysis in suggesting the various models which could explain the relation between observed variates. They cannot fully appreciate the use of this analysis until they themselves have formulated a model, and used the result of the regression calculation under this model to formulate another. They will learn from this experience, because passing through these various stages involves a continual process of re-evaluating the data.

Beyond statistics

The benefits of using the package extend beyond the study of statistics: students are introduced to a

A.

B.

BEST COPY AVAILABLE

```
► SHOWATABLE

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

          |  COLUMN CLASS NUMBER
ROW NAME  |     1.     2.     3.
-------------------------------
    MALE  |   10.    22.    28.
  FEMALE  |   14.    13.    13.

**NAMES OF COLUMN CLASSES**

  1. ↔ 0↑39
  2. ↔ 40↑59
  3. ↔ 60↑100

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

► CHISQUARE

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

DISPLAY EXPECTED FREQUENCIES?
►YES

FREQUENCIES AS PREDICTED BY NULL HYPOTHESIS

          |  COLUMN CLASS NUMBER
ROW NAME  |     1.     2.     3.
-------------------------------
    MALE  |  14.40  21.00  24.60
  FEMALE  |   9.60  14.00  16.40

CHI-SQUARE IS 4.654955478 FOR 2 D.F.

THE CORRESPONDING P-VALUE IS 0.0975
ACCEPT THE NULL HYPOTHESIS FOR ANY
P-VALUE LESS THAN THE VALUE GIVEN ABOVE.

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑
```

```
► MALE2↑DSTAT MALE

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

     SAMPLE SIZE IS  60
   MEAN OF SAMPLE IS  58.95
     SAMPLE S.D. IS  21.08451516
  SAMPLE VARIANCE IS  444.5567797

DISPLAY FURTHER STATISTICS?
►NO

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

► FEMALE2↑DSTAT FEMALE

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

     SAMPLE SIZE IS  40
   MEAN OF SAMPLE IS  49.875
     SAMPLE S.D. IS  16.19304615
  SAMPLE VARIANCE IS  262.2147436

DISPLAY FURTHER STATISTICS?
►NO

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

► HYCO MALE2 AND FEMALE2

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑

ENTER NUMBER OF DESIRED OPTION.

  1) TEST HYPOTHESIS ON MEAN
  2) ESTIMATE C.I. OF MEAN
  3) TEST HYPOTHESIS ON VARIANCE
  4) ESTIMATE C.I. OF VARIANCE
0:
►    1

ENTER DESIRED P-VALUE (0-1).
0:
►    0.05

IS POPULATION VARIANCE, STANDARD DEV., OR NEITHER KNOWN?
►N

THE COMPUTED TEST STATISTIC IS T=2.305076806 WITH 98 D.F.

THE PROBABILITY OF OBTAINING A T-VALUE ABOVE
THIS T IS APPROXIMATELY 0.0116 .

TAILS (1 OR 2)?
0:
►    1

REJECT NULL HYPOTHESIS AT P=0.05 .

↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑
```

FIGURE ONE. Either of these analyses could be performed on the examination scores of 60 males and 40 females to determine whether the males "demonstrated their superiority". The analyses differ in their assumptions, and give different conclusions using level of significance, p=0.05. In each analysis one begins by entering the data. This entry is not displayed in the figure. Lines marked with the symbol ► were typed by the student; all others were typed by the computer.

tool that has a growing influence in academic and everyday life.

Classroom experience with the package

The development in 1973 of the Cornell statistical programs [2] was supported by a National Fund for Medical Education grant to develop computer-related innovations for the education of medical students. The package was used in two biostatistics courses given at the Cornell Medical College in 1973 [3] and in a course on applications of computer science given at Cornell's School of Industrial Engineering in 1974. [4]

The biostatistics courses

In these courses students attended a 1/2 hour meeting in the room containing the computer terminals they would use. This meeting included a brief demonstration, distribution of the user's manual, and information on logistics of computer access. Prior to the courses, Professor Miké, her teaching assistant Ms. Mùshinski, and nearly all the students, had no exposure to the statistics package, APL language, or the College's interactive computing facilities. After the courses were over, Ms. Mushinski reported:

> Questions and problems were restricted to a few misinterpretations of the instructions in the manual and general confusion regarding the terminal keyboard. In every instance, the problems were easily corrected and the students continued to use to terminals. [5]

The package is intended to be a source of motivation, not of distaste. Yet no matter how excellent the package, an unreliable computer, or inaccessible computer terminals, will drive away most students. The Medical College, using the SECOS Computer Timesharing System,[6] fortunately had neither problem. And despite the fact that the students were busy and apprehensive, and their use of the package optional, their response was enthusiastic. Ms. Mushinski writes:

> Enthusiasm remained high as evidenced by the numbers of students who made use of the terminals and the APL Interactive Statistics Package as indicated below:
>
> -- Of the 65 students in the required Introduction to Biostatistics course, approximately 50 or more than 3/4 of the class participated in the experiment. The Statistics Package was used with very little difficulty and primarily in relation to the required homework problems. These students worked alone and in groups at the terminals. A few of the students indicated a desire to use the terminals on "more practical" or "more relevant" problems and they anticipated returning for such use later.
>
> -- Of the students in the elective course, approximately 10 or roughly one-third of the class used the APL Package on a fairly regular basis. These students were very enthusiastic about having access to the terminals and used them both in relation to their jobs and the weekly homework assignments. It is anticipated that a number of the staff members and a few of the medical and graduate students will continue to make use of the manual and the system as their needs require.

The applications of computer science course

Students in the course were assigned two problem sets to be worked using the package. They were not given instruction in using the package nor any demonstrations. They learned to use it solely from reading the manual and from trial and error. Although the teacher of the course, Professor Severance, and the computer personnel at the "public terminals" knew nothing about the package, they could and did assist the students in establishing connection to the computer.

After eight weeks of the course, each student was asked to write an evaluation of the package, along with suggestions for how it might be improved. By and large, the students thought well of their experience [7]:

> The manual is, for the most part, quite clear. The abundance of examples serves to clarify many possible hazy points.
>
> Most questions of syntax can be dealt with during actual use. The immediate response of the system also aids in learning what the programs can do....
>
> After comparing the length of time it took to do the first problem set [by hand] compared with the second [using the package], no one in our group had any complaints with the statistics package.

The students raised some problems:

> It is very difficult for groups of 5 to huddle around the APL terminal.
>
> The researcher who desires extremely flexible or specialized analysis is not going to be able to make extensive use of the package.

Only one problem mentioned by the students arose from the medical nature of the package design:

> ...there is some type of p-value given and no explanation of whether it's α, $1-\alpha$, or β.

The students' most common criticism was that the conversational nature of the package, which made it easy for beginners to use the computer, became a burden once beginners gained experience. Many students suggested the same remedy for this. For example:

> Perhaps Mr. Diegert could devise some system of making the amount of supportive information printed out by the computer depend on the experience of the user. Perhaps three categories would be appropriate: EXPERIENCED, SHAKY, AND INEXPERIENCED.

Students also suggested making the workspace bigger, tests of fit for more types of distributions, graphical display of regression output, covariance computation, and computation of more measures of a sample's central tendency.

Availability of the package

To use the package one must establish an APL billing account with a timesharing computer system that facilitates use of the APL computer language. In the engineering course, a unique account number was issued to each student, thus making individuals accountable for the computer charges incurred.

The package consists of statistical analysis procedures, viz. four "APL workspaces", that must be stored within the computer system. Usually this is done by the machine operators. These workspaces are normally stored in a "public

library" (as in the SECOS system), making them available to anyone with an APL account number using the system. Requests to export these workspaces to other computer installations may be made to:

Center for Environmental
Quality Management
468 Hollister Hall
Cornell University
Ithaca, N.Y. 14850

Copies of the user's manual are also available, at a cost of one dollar per copy.

29 March 1974

FOOTNOTES

[1] Existing APL statistical packages require that students know the particular format of the input and output data for each routine, and be reasonably fluent in APL language. Cf. K.W. Smillie, "STATPACK 2: An APL Statistical Package," 2nd ed., Publication #17, Department of Computer Science, University of Alberta, 1969; and J. Prins, "Statistical Programs in APL", 4th ed., State University College, New-Paltz, New York, 1972. In many cases it is appropriate to ask students to meet these requirements. But the professor, the students, or both, are often not willing -- and perhaps should not be expected -- to learn to interact with the computer at this level of detail.

[2] Carl Diegert, "An APL Interactive Statistics Package", Center for Environmental Quality Management Reprint #1039, Cornell University, Ithaca, New York, 1973.

[3] "Biostatistics" -- two courses offered by the Department of Public Health of the Cornell University Medical College, New York City; second trimester, 1973; taught by Dr. Valerie Miké.

[4] "Applications of Computer Science in Industrial Engineering and Operations Research" -- a course offered by the Department of Operations Research at Cornell University, Ithaca, New York; spring semester, 1974; taught by Dr. Dennis Severance.

[5] Correspondence from M.H. Mushinski, Senior Research Assistant, Cornell Medical School, April 1973.

[6] Shared Educational Computer System, 50 Market Street, Poughkeepsie, New York.

[7] Student evaluations courtesy of Dr. Severance.